Data in Brief 59 (2025) 111363

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Maternal health risk factors dataset: Clinical parameters and insights from rural Bangladesh

Mayen Uddin Mojumdar, Dhiman Sarker, Md Assaduzzaman*, Hasin Arman Shifa, Md. Anisul Haque Sajeeb, Oahidul Islam, Md Shadikul Bari, Mohammad Jahangir Alam, Narayan Ranjan Chakraborty

Department of CSE, Daffodil International University, Dhaka, Bangladesh

ARTICLE INFO

Article history:
Received 8 November 2024
Revised 28 January 2025
Accepted 29 January 2025
Available online 4 February 2025

Dataset link: Maternal Health Risk Assessment Dataset (Original data)

Keywords:
Early signs of pregnancy
Maternal health
Maternal health dataset
Maternal factors
Pregnancy dataset Bangladesh

ABSTRACT

Pregnancy-related complications and their consequences pose significant public health challenges, particularly in rural and developing areas where healthcare resources are limited. Monitoring clinical parameters during pregnancy improves diagnosis, treatment, and maternal health prognosis. This database includes records of pregnant patients from Kurigram General Hospital, Bangladesh. It captures core health parameters such as age, blood pressure (systolic and diastolic), blood sugar levels, body temperature, BMI, current mental health status, pre-existing medical history, gestational diabetes status, and heart rate. The diversity of data collected in this dataset is essential for understanding potential health changes associated with pregnancy. It will aid in generating high-risk pregnancy evaluation and prediction models to support clinical management. This dataset is valuable for its potential to serve as a benchmark for comparing maternal health responses across different clinical conditions of patients, thereby contributing to a broader understanding of pregnancy-related complications. The study's preprocessing methods, which included data cleaning, normalization, and encoding, ensured high-quality data for statistical analysis. Initial findings used statistical tests to explore associations within the data. A Chi-Square test analyzed the relationship

* Corresponding author.
E-mail address: assaduzzaman.cse@diu.edu.bd (M. Assaduzzaman).

https://doi.org/10.1016/j.dib.2025.111363

between preexisting diabetes and risk levels, revealing a significant association with a p-value of 4.85e-119. A Z-test was also conducted to compare clinical parameters between pregnant patients with and without diabetes, with a sample ratio of 337:811. This test showed a significant difference in BMI (body mass index), with a p-value of 2.23e-24, indicating that preexisting diabetes impacts BMI. A T-test for BMI revealed a significant difference, with a p-value of 1.405e-20. These findings further elucidate how specific age and body mass index details influence the risk levels associated with maternal clinical conditions. In summary, this database will be highly valued and a significant asset for research studies on maternal health in pregnant patients, public health strategies, and the enhancing diagnostic and treatment modalities for patients.

Specifications Table

Subject	Computer Science.
Specific subject area	Public health, Epidemiology, Clinical Health Indicators, Prediction, diagnosis, and treatment of pregnancy, Health Monitoring pregnancy
Type of data	Table, Dataset.
Data collection	Data was collected from the clinical parameters of pregnant patients at General Hospital, Kurigram, Bangladesh, on August 29, 2023, adhering strictly to ethical guidelines to ensure patient confidentiality of information. Patients' clinical profiles and vital signs were recorded during the initial diagnosis and follow-up visits to monitor health conditions and pregnancy-related risk factors. The dataset comprises key maternal clinical measurements, including Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, BMI (Body Mass Index), Previous Complications, Pre-existing Diabetes, Gestational Diabetes, Mental Health, and Heart Rate. This extensive collection of maternal clinical data has been anonymized and thoroughly cleaned to maintain accuracy and reliability. The dataset is instrumental in developing predictive models for early diagnosis, prognosis, and treatment strategies. By examining how different treatment approaches affect clinical outcomes, this data aims to support clinical decision-making and improve the management of pregnant patients.
Data source location	Town/City/Region: General Hospital, Kurigram. Country: Bangladesh.
Data accessibility	Repository name: Mendeley Data. Data identification number: 10.17632/p5w98dvbbk.1 Direct URL to data: https://data.mendeley.com/datasets/p5w98dvbbk/1 1
Related research article	None.

1. Value of the Data

- The dataset provides a comprehensive view of maternal health indicators, including blood pressure, BMI, blood sugar, and heart rate. These clinical parameters play a key role in assessing maternal health and identifying individuals at higher risk for pregnancy complications. The dataset includes all medically relevant aspects, such as age, systolic and diastolic blood pressure, blood sugar levels, body temperature, BMI, and heart rate. Additional condition indicators, like diabetes status and mental health, offer a thorough perspective on individual's physical and psychological well-being. This dataset is highly valuable for monitoring health status and assessing health risks within the maternal population.

- Machine learning models that predict pregnancy-related health risks may be trained using datasets, which are collections of maternal health markers. The analyst may use this data to create precise models without processing overhead. By applying transfer learning techniques and pre-trained models, researchers can improve model performance and reduce training time by using features learned from other health-related datasets. Predicting and categorizing risk levels is a key strength of models like Random Forest, XGBoost, and TabNet, which aid in early intervention initiatives. This approach can enhance the accuracy and efficiency of models designed to identify risk levels and detect early warning signs for maternal health complications, ultimately contributing to better patient outcomes.
- The dataset is a comprehensive resource for examining the effects of age, pre-existing conditions, and body metrics on individual health risks. It is uniquely suited for tracking trends and identifying health inequalities based on age and patient characteristics. Within the population, the dataset enables comparisons of how age impacts health risks and illnesses in conjunction with factors like pre-existing conditions, demographic characteristics, and specific body metrics such as Body Mass Index (BMI). Furthermore, the inclusion of data on both pre-existing diabetes and BMI clarifies diabetes-related health risks during pregnancy, thereby supporting research into maternal complications associated with diabetes. This dataset is also valuable for customized studies on risk factors, allowing researchers to explore relationships between health metrics—such as mental health status and heart rate—and maternal risk levels, ultimately contributing to the developing targeted intervention strategies.
- This dataset is valuable for testing and optimizing various healthcare strategies, including risk assessment, treatment guidelines, and resource allocation. It facilitates multidisciplinary research that combines clinical, physiological, and demographic data to aid in developing personalized healthcare approaches that improve patient outcomes. In practical applications, these data-driven solutions enhance healthcare providers' ability to accurately identify high-risk factors, ultimately leading to better health outcomes for patients, especially expectant mothers. The dataset supports educational initiatives for students and researchers and informs public health policies and maternal health programs

2. Background

Especially in rural and low-resource areas, pregnancy is a vital public health concern because of limited access to quality healthcare, inadequate nutrition, and insufficient prenatal and postnatal care. These areas face higher rates of maternal and infant mortality due to a shortage of skilled birth attendants and essential healthcare resources. The World Health Organization reports that millions of women worldwide face pregnancy-related risks, with nearly 800 deaths daily in 2020. The maternal mortality ratio dropped by 34% between 2000 and 2020 [1,2], with 95% in low and middle-income countries. Skilled health professionals can save lives and prevent complication. Diabetes, hypertension, and mental health problems are considered pregnancy complications that require close monitoring to avoid fatal outcomes for both the fetus and the mother [3]. The complications can lead to mortal conditions like preeclampsia, eclampsia, or maternal hemorrhage [4], which can raise the grade of maternal risk. Such complications are more related to the change in clinical parameters like blood pressure, BMI, mental health status, and pre-existing diabetes. It allows health professionals to further diagnose and treat high-risk pregnancies more professional through comprehensive monitoring.

Unmanaged, these complications can lead to life-threatening conditions such as preeclampsia, eclampsia, and maternal hemorrhage, which are major causes of maternal and perinatal mortality. Monitoring key clinical parameters like blood pressure, BMI, mental health status, and glucose levels is essential for early detection [5]. Mental health conditions like depression and anxiety can affect both the mother and baby, requiring support and possibly medication. Diabetes, especially gestational diabetes, needs blood sugar control to prevent complications like high birth weight or preterm birth. Hypertension can lead to serious risks like preeclampsia or restricted blood flow to the placenta. These issues are significant pregnancy complications that require consistent monitoring to prevent severe risks to both the mother and fetus [6]. Advanced diagnostic tools, wearable technologies, and mental health screening enhance real-time tracking and timely interventions. Comprehensive care and advanced technologies are crucial for reducing risks and improving outcomes in high-risk pregnancies [7].

Clinical criteria are crucial for monitoring maternal health, and comprehensive datasets that provide detailed health profiles of pregnant women are essential. Such data can enhance our understanding of pregnancy-related issues, improve diagnostic accuracy [8], and support the development of predictive models. This study aims to address existing gaps by providing an extensive dataset collected from pregnant women receiving care at the General Hospital in Kurigram, Bangladesh. The dataset includes age, systolic and diastolic blood pressure, blood sugar levels, body temperature, BMI, mental health status, medical history, preexisting and gestational diabetes, and heart rate readings. From this perspective, the wide range of parameters enables an in-depth analysis of clinical changes throughout pregnancy, making it valuable for assessing and managing high-risk pregnancies. This dataset can support the development of predictive models to improve diagnostic accuracy and enhance outcomes for pregnant patients [9]. It also has significant potential for medical and epidemiological studies, allowing for comparisons of maternal health responses across various demographics and regions [10].

Additionally, the dataset provides valuable insights for public health strategies and policies related to resource allocation and healthcare planning in maternal health management [11]. The knowledge derived from this data contributes not only to more refined clinical practices. It establishes a foundation for future studies in maternal and public health, ultimately supporting safer pregnancy management and improved maternal and child health outcomes.

3. Data Description

The dataset is a collection of clinical parameters of pregnant patients from at General Hospital, Kurigram, Bangladesh. It provides an overview of each individual's health risk level and demographics. The key considered vital signs were systolic and diastolic blood pressure, blood sugar level, body temperature, BMI, previous complications, pre-existing diabetes, and gestational diabetes. The other categorical outcomes included in the dataset are mental health, heart rate, and risk level. Age, blood sugar level, body temperature, BMI, previous complications, mental health status, heart rate, and risk level are all valuable in the analysis of health risk levels among patients. The dataset benefits health professionals in understanding and predicting Maternal health risk factor among patients. Table 1 provides a comprehensive overview of the key clinical parameters of the pregnant patients.

The following table presents a comprehensive explanation of each characteristic included in the dataset (Table 2).

4. Experimental Design, Materials and Methods

Age, Systolic Blood Pressure (BP), Diastolic Blood Pressure, Blood Sugar (BS), Body Temperature, BMI (Body Mass Index), Heart Rate, and Risk Level are key variables that correspond to the clinical parameters of pregnant patients which is visualizing at Fig. 1. Data preparation might involve several procedures, such as normalizing quantitative variables, encoding categorical variables, cleaning the dataset to address missing values, and imputing missing data. Errors may occur after all parameter-based data has been gathered and compiled, creating noise that could impair model testing performance and compromise the dataset's reliability. To address this, the dataset must cleaned using appropriate procedures to make it more dependable and noise-free. The histograms highlight key patterns, like Age and BMI, that are right-skewed, reflecting a higher number of younger individuals and those with healthy BMIs. Systolic and diastolic blood pressures, along with heart rates show nearly normal distributions with a slight right skew, meaning most people fall within typical ranges. Additionally, most participants have not experienced previous complications or gestational diabetes, as these variables are heavily

Table 1
Dataset overview of Maternal health risk factor dataset.

Age	Systolic BP	Diastolic	BS	Body Temp	BMI	PC	Pre-existing Diabetes	Gestational Diabetes	MH	Heart Rate	Risk Level
22	90	60	9	100	18	1	1	0	1	80	High
22	110	70	7.1	98	20.4	0	0	0	0	74	Low
27	110	70	7.5	98	23	1	0	0	0	72	Low
20	100	70	7.2	98	21.2	0	0	0	0	74	Low
20	90	60	7.5	98	19.7	0	0	0	0	74	Low
22	120	70	7.01	98	24	0	0	0	0	76	Low
20	110	70	9	102	17.6	0	1	0	0	78	High
23	110	80	7	98	21.3	0	0	0	0	74	Low
22	90	60	6.4	98	22	0	0	0	0	72	Low
26	110	70	12	100	30.2	1	1	1	1	80	High
25	100	60	7	98	24.5	0	1	0	0	74	Low
22	120	70	9.9	101	30	1	0	1	0	84	High
23	100	60	6.4	98	21	0	0	0	0	66	Low
19	90	60	6	98	21.5	0	0	0	0	70	Low
23	90	60	7	98	21	1	0	0	0	72	Low
18	100	70	7.5	98	18.6	0	1	0	0	74	Low
20	110	80	7.1	98	19	0	0	0	0	74	Low
24	110	60	6.5	98	21.2	0	0	0	0	70	Low
21	110	70	6.6	98	20	0	1	0	1	74	Low
44	140	100	14	98	25.3	0	1	0	1	88	High

Table 2
Feature description of dataset.

Feature	Description	Type	Unit/Value
Age	The patients' age, measured in years.	Numeric	Years
Systolic BP	Systolic blood pressure, representing the pressure in arteries when the heart beats.	Numeric	mmHg
Diastolic	Diastolic blood pressure, representing the pressure in arteries between beats.	Numeric	mmHg
BS	Blood sugar level, indicating the concentration of glucose in the blood.	Numeric	mmol/L
Body Temp	Body temperature of the patients.	Numeric	Degrees Fahrenheit (°F)
BMI	Body Mass Index, a measure of body fat based on weight and height.	Numeric	kg/m^2
Previous Complications	Indicates the presence of any past health complications (e.g., chronic conditions).	Binary	0,1
Pre-existing Diabetes	Shows whether the patient has a history of diabetes.	Binary	0,1
Gestational Diabetes	Indicates if the patient experienced gestational diabetes during pregnancy.	Binary	0,1
Mental Health	Binary indicator of mental health status.	Binary	0,1
Heart Rate	The number of heart beats per minute.	Numeric	Beats per minute (bpm)
Risk Level	The overall risk assessment level for the patient based on clinical parameters.	Categorical	Low, High

skewed toward zero. Other factors like body temperature, blood sugar, and mental health also show unique patterns.

4.1. Statistical analysis

The Chi-Square test is a statistical method used to see if there's a meaningful relationship between two categorical variables. It does this by comparing the actual data (observed frequencies) to what we would expect if there were no relationship between the variables (expected frequencies). The equ (i) test was used to analyze the relationship between preexisting diabetes

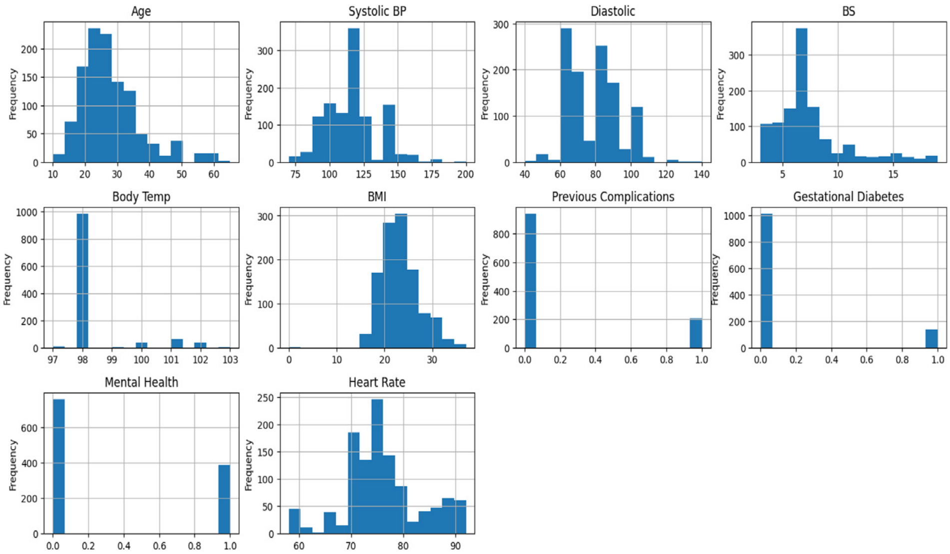

Fig. 1. Histograms of numeric features of parameters.

and risk levels, revealing a significant association with a p-value of 4.85e-119. A Z-test was also conducted to compare clinical parameters between pregnant patients with and without diabetes ratio 337:811. The Z-test showed a significant difference in BMI (Body Mass Index), with a p-value of 2.23e-24, indicating that preexisting diabetes impact BMI. Additionally, a T-test was performed to compare the means of each quantitative variable related to preexisting diabetes activity. A significant difference with a p-value of 1.405e-20 showed the T-test for Body Mass Index.

$$\mathrm{T} = \frac{X_1 - X_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{i}$$

Observing the equation, X_1 and X_2 are the sample means, where S_p represents the pooled standard deviation, and n_1 and n_2 are the sample sizes for each group. The differences identified in both the Z-test and the T-test revealed notable variations in body mass index based on the preexisting diabetes statistics, indicating that these statistics may influence this particular BMI parameter.

A one-way statistical method ANOVA was performed to compare the body mass index across different clinical outcomes. Following a boxplot, Fig. 2 visualizes the distribution of body mass index for each risk level related to the clinical outcome.

The IQR, median, and spread of data points for each clinical outcome of the risk levels (high and low) are represented in the figure, which visualizes Body Mass Index (BMI) across the final outcomes (risk levels). ANOVA testing was conducted to evaluate significant differences in BMI between the groups. Although the box plot shows some overlapping distributions, the ANOVA test did not reveal any statistically significant differences between the groups.

Fig. 3 displays a sample population's relationship between age and Body Mass Index (BMI). Each data point represents an individual, with age plotted on the x-axis and BMI on the y-axis. The scatterplot reveals a complex pattern. A slight upward trend in BMI is observed among younger individuals, likely due to natural growth and development. This trend stabilizes or decreases in early adulthood, possibly due to increased physical activity and a higher metabolic rate. However, in middle and older age, BMI tends to grow again, potentially due to reduced physical activity, hormonal changes, and a slower metabolism.

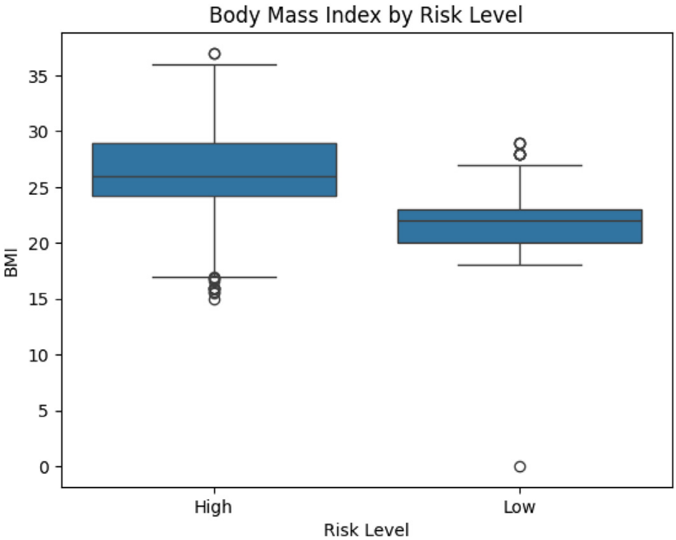

Fig. 2. Boxplot of Body Mass Index by risk level.

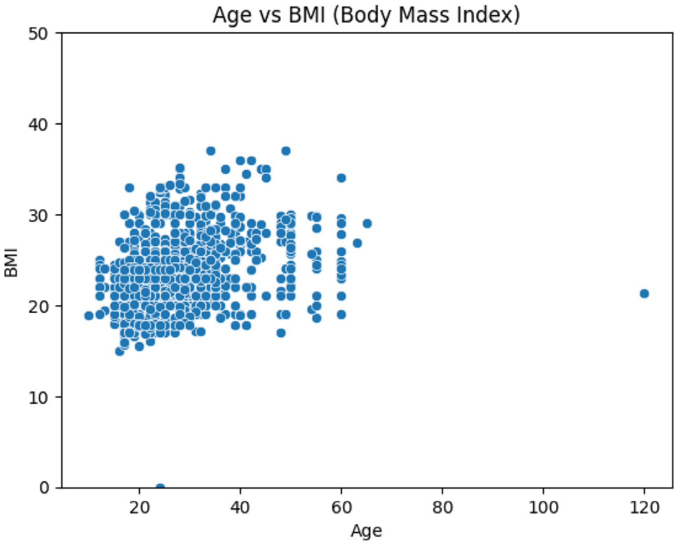

Fig. 3. Scatterplot of age vs BMI.

The scatter plot shows the relationship between the age and Body Mass Index (BMI) of pregnant patients, with each individual data point represented. The dataset highlights how the level of BMI varies with age. Based on the graph, a specific age group of patients appears to have some correlation with Body Mass Index. Fig. 4 displays a heatmap of the contingency table for preexisting diabetes versus risk levels as the final output, visualizing the distribution of outcomes across diabetes active status. The heatmap uses color intensity to visualize the frequency of individuals within each combination of diabetes status (Yes/No) and risk level (High/Low). The color indicates the higher proportions of individuals in each category. The heatmap reveals a strong association between preexisting diabetes and risk level. Individuals with diabetes are

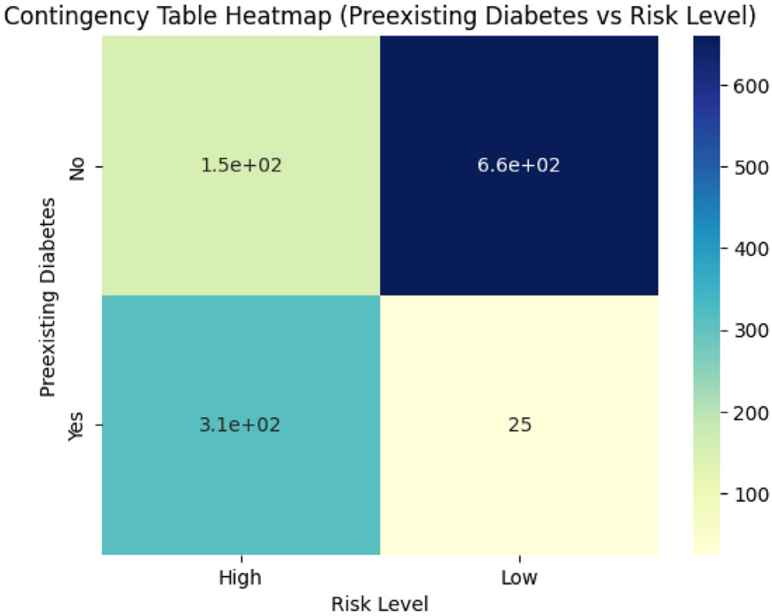

Fig. 4. Heatmap of the contingency table pre-existing diabetes vs risk level.

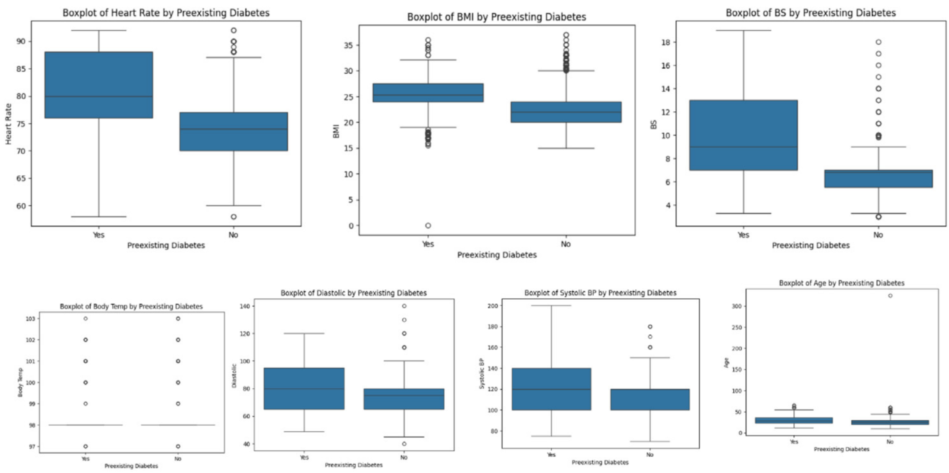

Fig. 5. Boxplots of clinical parameters of patients by pre-existing diabetes level.

disproportionately represented in the high-risk category (310 cases) compared to those without diabetes (150 cases). Conversely, the majority of individuals without diabetes fall into the low-risk category (660 cases), while only 25 individuals with diabetes are in this category. This disparity is statistically significant, as confirmed by the Chi-Square test.

The number of individuals with preexisting diabetes is substantially higher in the “High Risk” category compared to the “Low Risk” category, suggesting that individuals with diabetes are more likely to fall into the high-risk group. On the other hand, those without diabetes are predominantly concentrated in the low-risk category. The heatmap clearly depicts the distribution of various outcomes across different diabetes statuses.

Table 3
Statistical analysis of test result.

Test	Variable	P-value
Z-Test	Pre-existing Diabetes vs Risk Level	4.85e-119
Z-Test	Age	6.87e-09
Z-Test	Blood Pressure	8.76e-09
Z-Test	Diastolic	3.67e-11
Z-Test	Blood Sugar	4.51e-112
Z-Test	Body Temperature	0.013
Z-Test	Body Mass Index	2.22e-24
Z-Test	Heart Rate	9.631e-42
T-Test	Age	2.42e-10
T-Test	Blood Pressure	1.86e-07
T-Test	Diastolic	8.49e-10
T-Test	Blood Sugar	1.078e-48
T-Test	Body Temperature	0.020
T-Test	Body Mass Index	1.40e-20
T-Test	Heart Rate	5.66e-27

Finally, Fig. 5 shows the differences in clinical parameters of pregnant patients with active and inactive preexisting diabetes by displaying boxplots for quantitative variables based on diabetes status. The distribution of several clinical parameters among individuals with preexisting diabetes, both active and inactive, is shown in these boxplots. Each boxplot displays the median, interquartile range (IQR), and potential outliers for each parameter. The Z-test and T-test for Body Mass Index were used to compare the means of these parameters. With a p-value of 1.405e-20, the T-test for Body Mass Index levels revealed a significant difference between active and inactive diabetes, indicating that preexisting diabetes influences BMI levels.

Table 3, which provides the p-values for each test conducted on the clinical parameters of pregnant women, summarizes the findings from the various statistical tests. These studies illustrate the importance of considering preexisting diabetes status in medical diagnostics by revealing significant changes in certain clinical indicators, such as body mass index.

Together, these analyses and visualizations enhance our understanding of how changes in BMI relate to the risk levels of pregnant patients, offering crucial new information for improving diagnostic and treatment approaches.

Limitations

The dataset provides valuable insights from General Hospital, Kurigram, Bangladesh, but validation in other areas could enhance its generalizability for studies on clinical health reports of pregnant patients. While larger datasets might yield more robust results, the 1,206 entries represent a diverse range of clinical responses. Although the study focused on specified key clinical parameters of pregnant patients, future research could gain deeper insights by incorporating additional factors such as co-morbidities.

Ethics Statement

The data collection adhered to strict ethical guidelines, with informed consent obtained from all participants or their legal guardians. Patient confidentiality was safeguarded through data anonymization. Ethical approval was granted by the review board in General Hospital, Kurigram, Bangladesh, on August 29, 2023. All data were used solely for research purposes, following both local and international ethical standards.

Credit Author Statement

Mayen Uddin Mojumdar: Conceptualization, Methodology, Data Curation, Supervision. **Dhiman Sarker:** Software, Writing Original Draft. **Md Assaduzzaman:** Supervision, Methodology, Project administration. **Hasin Arman Shifa:** Data Curation. **Md. Anisul Haque Sajeeb:** Visualization. **Oahidul Islam:** Project administration. **Md Shadikul Bari:** Resources. **Mohammad Jahangir Alam:** Methodology. **Narayan Ranjan Chakraborty:** Writing - Review.

Data Availability

Maternal Health Risk Assessment Dataset (Original data) (Mendeley Data).

Acknowledgements

We extend our sincere gratitude to Dr. Shahinur Rahman Sardar, Residential Medical Officer and domain expert at General Hospital, Kurigram, for her valuable insights and support, which were instrumental in the completion of this paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.T. Hossain, et al., Effective multi-sectoral approach for rapid reduction in maternal and neonatal mortality: the exceptional case of Bangladesh, BMJ Glob. Health 9 (Suppl 2) (2024) e011407 May, doi:10.1136/bmjgh-2022-011407.
[2] World Health Organization, “Maternal mortality ratio (modeled estimate, per 100,000 live births),” 2020. [Online]. Available: https://data.worldbank.org/indicator/SH.STA.MMRT. Accessed: 2020.
[3] M.A. Khan, A.L. Wojdyla, E. Say, et al., WHO analysis of causes of maternal death: a systematic review, Lancet 367 (9516) (2006) 1066–1074.
[4] M.W. Stewart, S.J. Graham, P.L. Appleby, et al., Prevalence and predictors of pre-eclampsia in pregnant women with pre-existing diabetes, Hypertens. Pregnancy 25 (1) (2015) 57–66.
[5] J.L. Horton, C. Levin, Diabetes and pregnancy: An international comparison of diabetes and related health care services, Diabetes Care 26 (2003) 2386–2391.
[6] M. Assaduzzaman, A.A. Mamun, M.Z. Hasan, Early prediction of maternal health risk factors using machine learning techniques, in: 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1–6, doi:10.1109/ICONAT57137.2023.10080700.
[7] L Jamel, M Umer, O Saidani, B Alabduallah, S Alsubai, F Ishmanov, T Kim, I. Ashraf, Improving prediction of maternal health risks using PCA features and TreeNet model, PeerJ. Comput. Sci. 10 (2024) e1982, doi:10.7717/peerj-cs.1982.
[8] S.F. Lee, The importance of skilled health professionals in reducing maternal and infant mortality, Int. J. Public Health 5 (2) (2021) 56–65.
[9] S.A. Saleem, J. Fikree, Maternal and infant health monitoring in low-resource settings, Public Health Rev. 36 (3) (2019) 1–9.
[10] T. Ross, Mental health and pregnancy: addressing gaps in prenatal and postnatal care, J. Public Health 27 (1) (2022) 112–121.
[11] H.Y. Kim, et al., Data analysis and monitoring of high-risk pregnancies, Health Inform. J. 25 (1) (2019) 73–85 Mar., doi:10.1177/1460458218757195.